

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Computational and Applied Mathematics 193 (2006) 140–156

JOURNAL OF
COMPUTATIONAL AND
APPLIED MATHEMATICSwww.elsevier.com/locate/cam

Plane wave decomposition in the unit disc: Convergence estimates and computational aspects

E. Perrey-Debain

School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK

Received 13 October 2004

Abstract

This paper deals with the numerical simulation of time-harmonic wave fields using progressive plane waves. It is shown that a plane wave travelling in arbitrary direction can be numerically recovered with an accuracy of the order of the machine precision with a collocation formulation and the square root of the machine precision with a least-square formulation. However, strongly evanescent and nearly singular wave fields cannot be properly recovered with standard double-precision floating-point arithmetic. Some of the ideas are applied to the elastic wave equation and a simple optimization algorithm is proposed to find a good compromise between the accuracy and the number of plane waves.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Plane wave basis; Helmholtz equation; Elastodynamics; Finite precision computation

1. Introduction

Methods using superposition of progressive plane waves for the numerical simulation of time-harmonic wave problems generally falls in the much wider class of methods called Trefftz-type methods in which an approximate solution of a boundary value problem is built from the sets of functions that satisfy exactly the differential equation. These plane wave methods have been mainly developed for domain discretization schemes. Although this is not the place for a complete survey, one can cite the Ultra Weak Formulation introduced by Després for the Helmholtz equation [7,6,9] and recently extended for the elastodynamic equation [8] or the least-squares Trefftz-type elements [11]. Use of plane waves is also

E-mail address: emmanual@maths.man.ac.uk.

advocated in the Partition of Unity Method introduced by Babuška and Melenk [2] and applications for scattering problems can be found in [10,13,12]. All these techniques showed considerable improvements both in terms of degree of freedom reduction and accuracy compared with conventional discretization schemes. However, the question of numerical stability of the plane wave basis due to the poor conditioning of the resulting algebraic system remains an open problem. Sometimes described as basis ‘badness’ in quantum mechanics [3], this can bring severe limitations to the method if the wave field to be approximated is strongly evanescent. Though evanescent waves can theoretically be expressed as the singular limit of an angular superposition of real (i.e. progressive) plane waves [4], their associated coefficients become exponentially large so that only many-decimal arithmetic computation can recover the exact solution.

The present paper aims at bringing some new contributions to the understanding of these matters. Focusing on the Helmholtz equation in the unit disc, precise estimates for the plane wave basis approximation error (in the maximum-norm) as well as the conditioning number arising from both least square and collocation formulations are given in Section 2. In Section 3, some of the ideas developed for the Helmholtz problem are applied to the elastic wave equation.

2. Helmholtz equation

In this section, we consider the Helmholtz equation on a circular domain of diameter h . Without lack of generality we restrict ourselves to the particular case where the domain Ω is the unit disc by introducing the reduced wave number $\kappa = \pi h / \lambda$ (λ is the wavelength) so that the Dirichlet problem can be written as

$$\Delta u + \kappa^2 u = 0 \quad \text{on } \Omega, \quad (1)$$

$$u = g \quad \text{on } \gamma = \partial\Omega. \quad (2)$$

In the sequel, we call $\mathbf{x} = (x_1, x_2)$ the cartesian coordinates and (r, θ) , its polar representation. We note $\langle \cdot, \cdot \rangle$ and $\| \cdot \|_{L^2(\gamma)}$ the usual inner product and its associated norm of the Hilbert space $L^2(\gamma)$.

2.1. Error analysis

We assume that the boundary data g are given via its Fourier series as

$$g(\theta) = \sum_{n \in \mathbb{Z}} \hat{g}_n e^{in\theta}, \quad (3)$$

where the series converges pointwise on $[0, 2\pi]$. Provided that the wave number κ is such that $J_n(\kappa) \neq 0$ for any integer $|n| < \kappa$, the unique solution is given by the infinite sum

$$u(\mathbf{x}) = \sum_{n \in \mathbb{Z}} \hat{g}_n \frac{J_n(\kappa r)}{J_n(\kappa)} e^{in\theta}. \quad (4)$$

We define by u_N the truncated sum (4) up to the order N and we call

$$\Psi(\kappa; \phi, \mathbf{x}) = \exp(i\kappa(x_1 \cos \phi + x_2 \sin \phi)) \quad (5)$$

a progressive plane wave in the direction ϕ . By using Bessel's first integral identity [1], u_N can be expanded with plane wave integrals as

$$u_N(\mathbf{x}) = \sum_{|n| \leq N} \frac{\hat{g}_n}{J_n(\kappa)} \frac{1}{2\pi i^n} \int_0^{2\pi} \Psi(\kappa; \phi, \mathbf{x}) e^{in\phi} d\phi. \quad (6)$$

Evaluating integrals in (6) with the trapezoidal rule using a fixed set of quadrature points $\phi_q = 2\pi q/Q$ yields the plane wave approximation

$$\tilde{u}_{Q,N}(\mathbf{x}) = \sum_{q=1}^Q \Psi_q(\kappa; \mathbf{x}) \left(\frac{1}{Q} \sum_{|n| \leq N} \frac{\hat{g}_n e^{in2\pi q/Q}}{J_n(\kappa) i^n} \right), \quad (7)$$

where the set Ψ_q are progressive plane waves travelling in directions evenly distributed over the unit circle, $\Psi_q(\kappa; \mathbf{x}) = \Psi(\kappa; \phi_q, \mathbf{x})$ for $q = 1, 2, \dots, Q$. In order to give an estimation of the approximation error $u_N - \tilde{u}_{Q,N}$, call $\varepsilon_{Q,n}$ the quadrature error

$$\varepsilon_{Q,n}(\kappa; \mathbf{x}) = \int_0^{2\pi} \Psi(\kappa; \phi, \mathbf{x}) e^{in\phi} d\phi - \frac{2\pi}{Q} \sum_{q=1}^Q \Psi_q(\kappa; \mathbf{x}) e^{in2\pi q/Q}. \quad (8)$$

Using the Jacobi–Anger expansion for the plane wave [1], we get

$$\begin{aligned} \varepsilon_{Q,n}(\kappa; \mathbf{x}) &= \sum_{m \in \mathbb{Z}} i^m J_m(\kappa r) e^{im\theta} \left(\int_0^{2\pi} e^{i(n-m)\phi} d\phi - \frac{2\pi}{Q} \sum_{q=1}^Q e^{i(n-m)2\pi q/Q} \right) \\ &= 2\pi \sum_{m \in \mathbb{Z}} i^m J_m(\kappa r) e^{im\theta} \left(\delta_{n,m} - \sum_{k \in \mathbb{Z}} \delta_{n-m, kQ} \right) \\ &= -2\pi \sum_{k \in \mathbb{Z} \setminus \{0\}} i^{n+kQ} J_{n+kQ}(\kappa r) e^{i(n+kQ)\theta}, \end{aligned} \quad (9)$$

where δ is the Kronecker symbol. A upper bound for the norm of the quadrature error $\varepsilon_{Q,n}$ can be given if the number of plane waves Q exceeds $|n| + \kappa$. More precisely, we show in Appendix B that, if the number of plane waves is chosen such that $Q = N + \beta_N \kappa$ with $\beta_N > 1$ then the following inequalities hold:

$$\|\varepsilon_{Q,n}\|_{L^\infty(\Omega)} < \frac{4\pi\beta_N^Q}{\beta_N^Q - 1} J_{Q-N}(\kappa) \quad \forall |n| \leq N. \quad (10)$$

Thus, under the same conditions, we have

$$\begin{aligned} \|u_N - \tilde{u}_{Q,N}\|_{L^\infty(\Omega)} &\leq \frac{1}{2\pi} \sum_{|n| \leq N} \left| \frac{\hat{g}_n}{J_n(\kappa)} \right| \|\varepsilon_{Q,n}\|_{L^\infty(\Omega)} \\ &< \frac{2\beta_N^Q}{\beta_N^Q - 1} J_{Q-N}(\kappa) \sum_{|n| \leq N} \left| \frac{\hat{g}_n}{J_n(\kappa)} \right|. \end{aligned} \quad (11)$$

Now, using the properties of Bessel functions (see Appendix A), it is straightforward to see that provided $N \geq \kappa$ then

$$\|u - u_N\|_{L^\infty(\Omega)} \leq \sum_{|n| > N} |\hat{g}_n|. \quad (12)$$

By virtue of (11) and (12), we can now state the following lemma:

Lemma 1. *Define the system of plane waves*

$$W(Q) = \text{span}\{\Psi(\kappa; \phi_q, \mathbf{x}), \phi_q = 2\pi q/Q, q = 1, \dots, Q\}.$$

Let the number of plane waves $Q > 2\kappa$ be chosen such that the set $I = [\kappa, Q - \kappa] \cap \mathbb{N}$ is not empty. Then, the best approximation in $W(Q)$ of the Dirichlet problem (1), (2) satisfies the inequality

$$\min_{w \in W(Q)} \|u - w\|_{L^\infty(\Omega)} < \min_{N \in I} \left\{ \sum_{|n| > N} |\hat{g}_n| + \frac{2\beta_N^Q J_{Q-N}(\kappa)}{\beta_N^Q - 1} \sum_{|n| \leq N} \left| \frac{\hat{g}_n}{J_n(\kappa)} \right| \right\}, \quad (13)$$

where $\beta_N = (Q - N)/\kappa$.

More precise estimates can be derived if the Fourier series (3) is finite, i.e. $\hat{g}_n = 0$ for all $|n| > n_0$. In this case we have

Lemma 2. *Let the number of plane waves $Q > \kappa + n_0$. Then, the best approximation in $W(Q)$ of the Dirichlet problem (1), (2) satisfies the inequality*

$$\min_{w \in W(Q)} \|u - w\|_{L^\infty(\Omega)} < \frac{2\beta_{n_0}^Q J_{Q-n_0}(\kappa)}{\beta_{n_0}^Q - 1} \sum_{|n| \leq n_0} \left| \frac{\hat{g}_n}{J_n(\kappa)} \right|, \quad (14)$$

where $\beta_{n_0} = (Q - n_0)/\kappa$.

We shall apply these error estimates in two cases. First, assume the data g are given from a homogeneous plane wave travelling in the arbitrary direction ϕ , i.e.

$$u_I(\mathbf{x}) = \Psi(\kappa; \phi, \mathbf{x}). \quad (15)$$

In that case, we have $|\hat{g}_n| = |J_n(\kappa)|$. By following similar techniques as in Appendix B, it is easy to see that, if $N > \kappa$ then

$$\sum_{|n| > N} |\hat{g}_n| < \frac{2J_N(\kappa)}{N/\kappa - 1}.$$

Thus, a sharp estimate (13) is obtained if $N \in I$ is chosen such that $J_{Q-N}(\kappa)$ and $J_N(\kappa)$ are minimized simultaneously and $N = Q/2$ (Q is assumed even for simplicity) appears to be a reasonable choice. This yields the following estimate:

$$\min_{w \in W(Q)} \|u_I - w\|_{L^\infty(\Omega)} < 2J_{\beta\kappa}(\kappa) \left(\frac{1}{\beta - 1} + \frac{\beta^Q(Q + 1)}{\beta^Q - 1} \right), \quad (16)$$

where

$$\beta = Q/2\kappa > 1$$

(note that the quantity 2β can be interpreted as the number of degrees of freedom per full wavelength on the perimeter of γ). This last result shows that any progressive plane wave of unit amplitude can be approximated in $W(Q)$ with an error behaving like $QJ_{Q/2}(\kappa)$. Moreover, by virtue of (A.8), the error will be infinitely small in the *high-frequency limit* when $Q_{\text{HF}} = e\kappa$.

Remark. (i) From a practical point of view, the condition $Q > 2\kappa$ is not penalizing compared to conventional domain discretization schemes for which the number of variables needed to approximate the wave field in Ω behaves like $(\tau\kappa)^2/(4\pi)$. Here, the parameter τ stands for the discretization level which usually lies around 10 variables per wavelength. The fact that Q almost behaves linearly with κ was expected since only the boundary data are approximated (like any boundary integral method).

(ii) Further analysis could be carried out in the spirit of [5] to evaluate the number of plane waves Q_{\min} needed to guarantee an approximation error below a certain value (say ε). Since for large κ , the function $f(Q) = QJ_{Q/2}(\kappa)$, $Q > 2\kappa$ is very steep, it is anticipated that $Q_{\min} = 2\kappa + q(\varepsilon, \kappa)$. For a fixed constraint ε , the quantity 2κ becomes the dominant term as κ increases and this is very similar to the analysis given in [5, Remark 3.3. p. 390].

Now, assume that the data g are given by $\hat{g}_n = \delta_{n,n_0}$ ($n_0 \geq 0$) so that the exact solution is simply

$$u_{II}(x) = \frac{J_{n_0}(\kappa r)}{J_{n_0}(\kappa)} e^{in_0\theta}. \quad (17)$$

Let $Q > \kappa + n_0$, then the best plane wave approximation satisfies

$$\min_{w \in W(Q)} \|u_{II} - w\|_{L^\infty(\Omega)} < \frac{2\beta_{n_0}^Q}{\beta_{n_0}^Q - 1} \frac{J_{Q-n_0}(\kappa)}{|J_{n_0}(\kappa)|}. \quad (18)$$

2.2. The conditioning problem

Inspection of the plane wave coefficients in (7) reveals that their magnitudes are not bounded and they display unpredictable behavior. This suggests that any algorithm devised to find these coefficients will probably face ill-conditioning problems. Fortunately, when the computational domain is circular, progress can be made towards an understanding of this effect. We shall start with the least-square method.

2.2.1. Least-square method

We seek an approximate solution of (1) and (2) by considering the linear combination of plane waves

$$w(x) = \sum_{q=1}^Q a_q \Psi_q(\kappa; x) \quad (19)$$

and choose the coefficients a_q so as to minimize the L_2 norm error $E = \|w - g\|_{L^2(\gamma)}$. Standard calculations demonstrate that the vector $\mathbf{a} = (a_1, a_2, \dots, a_Q)^T$ minimizes E if and only if a_q satisfies the normal

equations

$$\sum_{q=1}^Q \langle \Psi_q, \Psi_p \rangle a_q = \langle g, \Psi_p \rangle, \quad p = 1, 2, \dots, Q. \quad (20)$$

The $Q \times Q$ matrix M with entries $m_{pq} = \langle \Psi_q, \Psi_p \rangle$ is obviously Hermitian and positive since by definition, $\mathbf{a}^H M \mathbf{a} = \|\mathbf{w}\|_{L^2(\gamma)}^2$. (The symbol H denotes the conjugate transpose.) M is invertible provided that the functions Ψ_q are linearly independent in γ . This is true as long as κ is not an eigenvalue of the interior Dirichlet problem (see [6]) and therefore, the minimization problem is well-defined and we can define

$$E_Q = \min_{w \in W(Q)} \|w - g\|_{L^2(\gamma)}. \quad (21)$$

Using the Jacobi–Anger expansion and the orthogonality of the Fourier basis, the element matrice m_{pq} can be decomposed as follows:

$$\begin{aligned} m_{pq} &= \int_{\gamma} \Psi(\kappa; \phi_q, \mathbf{x}) \overline{\Psi(\kappa; \phi_p, \mathbf{x})} d\gamma \\ &= 2\pi \sum_{m \in \mathbb{Z}} J_m^2(\kappa) e^{im(\phi_p - \phi_q)} \\ &= 2\pi \sum_{k=1}^Q \sum_{l \in \mathbb{Z}} J_{k+lQ}^2(\kappa) e^{i2\pi(k+lQ)(p-q)/Q} \\ &= 2\pi \sum_{k=1}^Q e^{i2\pi kp/Q} \left(\sum_{l \in \mathbb{Z}} J_{k+lQ}^2(\kappa) \right) e^{-i2\pi kq/Q}. \end{aligned}$$

Thus, if we call W the (unitary) Discrete Fourier Transform matrix with entries $w_{pk} = Q^{-1/2} e^{i2\pi kp/Q}$ we have the diagonalization of M as

$$M = 2\pi Q W \Sigma W^H, \quad (22)$$

where Σ is the diagonal matrix containing the singular values

$$\sigma_k = \sum_{l \in \mathbb{Z}} J_{k+lQ}^2(\kappa). \quad (23)$$

We show in Appendix B that, if $Q > 2\kappa$ (Q even for simplicity), then

$$\min_{1 \leq k \leq Q} \sigma_k < 5J_{Q/2}^2(\kappa). \quad (24)$$

This gives us a lower bound for the condition number (in the 2-norm) as

$$\text{cond}_2(M) = \|M\|_2 \cdot \|M^{-1}\|_2 > \frac{\max_{1 \leq k \leq Q} J_k^2(\kappa)}{5J_{Q/2}^2(\kappa)} \quad (Q > 2\kappa). \quad (25)$$

2.2.2. Collocation method

A diagonalization of the collocation matrix can be obtained analytically if the plane wave expansion (19) is applied at Q points regularly distributed over γ , $\mathbf{x}_p = (\cos(2\pi p/Q), \sin(2\pi p/Q))$. Following the same technique as in the previous discussion, we find that the collocation matrix $(C)_{pq} = \Psi_q(\kappa; \mathbf{x}_p)$ admits the factorization

$$C = QDW^H, \quad (26)$$

where D is a diagonal matrix with coefficients

$$d_k = \sum_{l \in \mathbb{Z}} i^{(k+lQ)} J_{k+lQ}(\kappa). \quad (27)$$

By considering the normal matrix $C^H C$, we obtain formally

$$\text{cond}_2(C) = \left(\frac{\max_{1 \leq k \leq Q} (d_k \bar{d}_k)}{\min_{1 \leq k \leq Q} (d_k \bar{d}_k)} \right)^{1/2}. \quad (28)$$

To make some progress, it suffices to observe that, by using the trapezoidal rule, the least-square matrix can be decomposed as follows:

$$M = \frac{2\pi}{Q} C^H C + R(Q), \quad (29)$$

where the residual matrix $R(Q)$ stems from the quadrature error of a smooth 2π -periodic function and therefore tends to zero exponentially fast. Thus,

$$|d_k| \approx \sigma_k^{1/2} \quad \text{and} \quad \text{cond}_2(C) \approx (\text{cond}_2(M))^{1/2} \quad \text{as } Q \rightarrow \infty. \quad (30)$$

2.3. Numerical experiments—the finite precision problem

In the first part of this section, we shall check the error estimates as well as the condition number arising from both least-square and collocation formulations. In all the calculations, we ensure that $\min_{|n| < \kappa} |J_n(\kappa)|$ is not too small and certainly much higher than the machine precision so that our analysis is not spoiled by the nonuniqueness problem.

Let the boundary data g stemming from a progressive plane wave travelling in the direction $\phi_Q = \pi/Q$. The exact solution is simply

$$u_I(\mathbf{x}) = \Psi(\kappa; \phi_Q, \mathbf{x}).$$

With this choice, we ensure that none of the plane waves $\Psi_{q=1, \dots, Q}$ coincides with, or is too close to, the original plane wave. Fig. 1 shows the evolution of the condition number for both formulations and for two values of the wave number. Theoretical estimates are computed from (25) and (30). Above a threshold estimated at 10^{16} , the computer fails in evaluating correctly the condition number (and more precisely the smallest singular values) because the machine precision is reached. The effect of this finite precision problem on the L_2 norm error \hat{E}_Q is shown in Fig. 2 (the hat symbol refers to the computed version of E_Q in (21)). Note that $E_Q \leq \sqrt{2\pi} \min_{w \in W(Q)} \|w - u_I\|_{L^\infty(\Omega)}$, so the theoretical estimate (16) has been multiplied by the factor $\sqrt{2\pi}$ to give a fair comparison with \hat{E}_Q . The collocation formulation benefits from

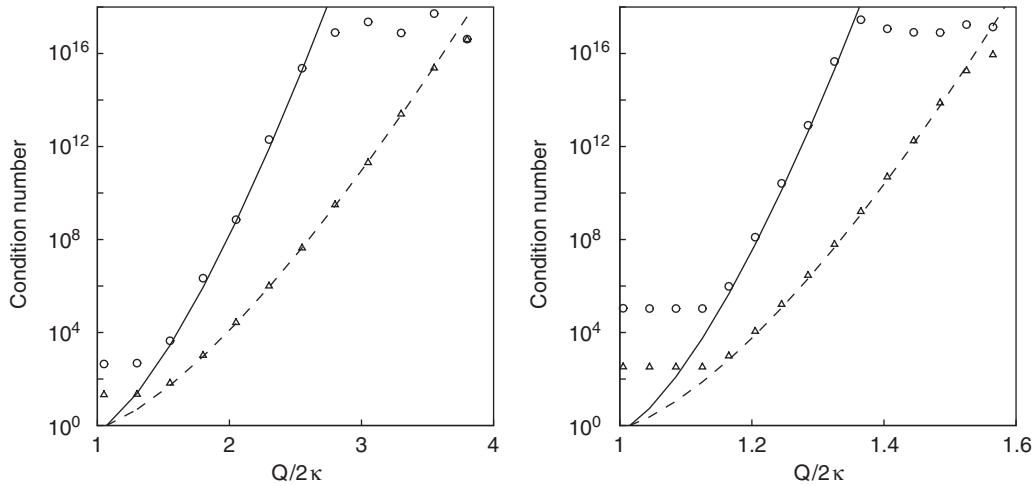


Fig. 1. Evolution of the condition number. Left: $\kappa = 10$. Right: $\kappa = 100$. Least-square formulation: theoretical estimate (straight line) and computed (circle). Collocation formulation: theoretical estimate (dashed line) and computed (delta).

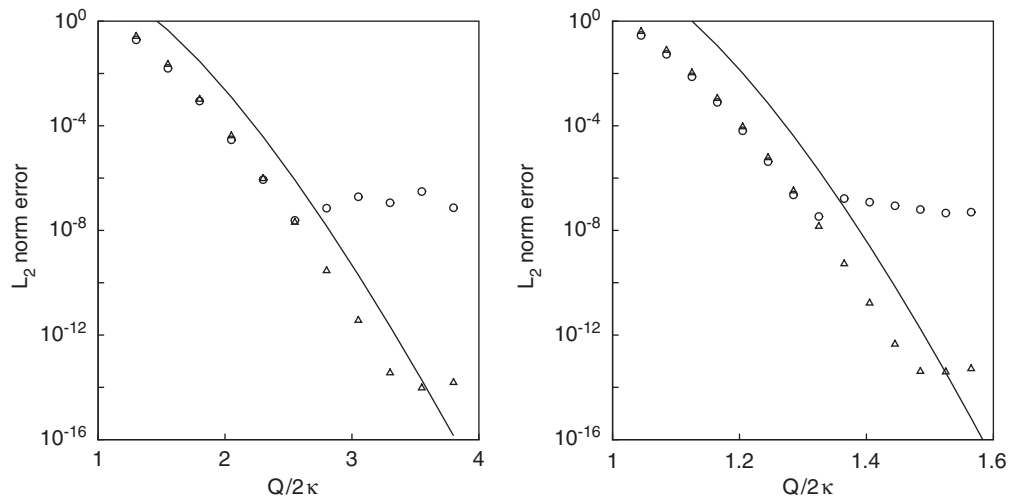


Fig. 2. Evolution of the error when approximating a plane wave. Left: $\kappa = 10$. Right: $\kappa = 100$. Theoretical estimate (straight line). Computed from collocation formulation (delta). Computed from least-square formulation (circle).

a better conditioning and gives the best results when $g(\theta)$ is infinitely differentiable, which is the case in this example. At $\kappa = 100$, the error behavior is close to the high-frequency regime ($Q_{\text{HF}}/(2\kappa) \approx 1.3591 \dots$). Though numerical results agree with our analysis, it appears that estimate (16) is overestimated and more accurate predictions could possibly be found for this particular problem. Nevertheless, this is not of crucial importance in practice, since for sufficiently large κ , $f(Q) = QJ_{Q/2}(\kappa)$ is a very steep function when $Q > 2\kappa$ and (16) remains a good indicator.

Finite precision calculations can have severe consequences in situations where g contains nonnegligible Fourier series coefficients in the range $|n| > \kappa$. Consider the collocation formulation for instance. Let us

first assume the ideal case where the inversion of the collocation matrix (26) can be achieved without loss of accuracy so that the plane waves coefficients are explicitly given by the convolution product

$$a = \frac{1}{Q} \mathbf{W} \mathbf{D}^{-1} \mathbf{W}^H \mathbf{g}, \quad (\mathbf{g})_p = g(2\pi p/Q). \quad (31)$$

Call \mathbf{W}_k the k th column of \mathbf{W} . Then we can rewrite (31) in terms of \hat{g} as

$$a = \frac{1}{\sqrt{Q}} \sum_{k=1}^Q b_k \mathbf{W}_k \quad \text{where } b_k = \frac{\sum_{l \in \mathbb{Z}} \hat{g}_{k+lQ}}{d_k}. \quad (32)$$

Now, let Q be arbitrary high so that coefficients amplitudes $|b_k|$ are fairly approximated by

$$|b_k| \approx \frac{|\sum_{l \in \mathbb{Z}} \hat{g}_{k+lQ}|}{\sigma_k^{1/2}}. \quad (33)$$

Introduce the index $k_\varepsilon > \kappa$ such that $J_{k_\varepsilon}(\kappa) < \varepsilon$, then by using the properties of the Bessel function and Annexe B.1, it is clear that

$$\sigma_k < 5\varepsilon^2 \quad \text{in the interval } k_\varepsilon < k < Q - k_\varepsilon. \quad (34)$$

Thus,

$$|b_k| > \frac{|\sum_{l \in \mathbb{Z}} \hat{g}_{k+lQ}|}{\sqrt{5\varepsilon}}. \quad (35)$$

In other words, high-order coefficients \hat{g}_n , ($|n| > k_\varepsilon > \kappa$) are magnified at least by the factor ε^{-1} . So if (32) as well as the plane wave expansion (19) are computed with finite precision, the information contained in these coefficients will be lost and only many-decimal arithmetic computation can circumvent this problem. In practice, the matrix inversion is carried out numerically (using the SVD algorithm in our case). With standard double-precision floating-point arithmetic, ε numerically stabilizes at the machine precision ($\sim 10^{-16}$) and the information is therefore already lost at that stage.

To illustrate this matter, let us consider approximating the exact solution u_{II} (see (17)) with plane waves using the collocation formulation. Since $\hat{g}_{n_0} = \delta_{n,n_0}$, this yields explicitly

$$a_q = \frac{e^{i2\pi q n_0/Q}}{Q d_{n_0}}. \quad (36)$$

When Q increases, d_{n_0} tends to $i^{n_0} J_{n_0}(\kappa)$ and the exact solution is recovered (see (7)). Nevertheless, if n_0 is chosen substantially above κ , then d_{n_0} tends very rapidly to zero like (A.8) and standard double precision quickly becomes inefficient. Tables 1 and 2 clearly shows the degradation of the error when $n_0 > \kappa$. The second line refers to the collocation formulation and the system is numerically inverted and the third line refers to the exact plane wave approximation with coefficients $a_q = e^{i2\pi q n_0/Q} / (Q i^{n_0} J_{n_0}(\kappa))$. The number of plane waves Q is taken high enough to ensure that errors have stabilized. It can be noticed that the finite precision effect is relatively more severe at high frequency. In Fig. 3, are plotted the computed plane wave approximations of (17) at $\kappa = 10$ and for the four values $n_0 = 10, 20, 25$ and 30 are plotted. The expected concentric circles become distorted until the information is completely lost when $n_0 > 30$.

Table 1

Degradation of the error due to the finite precision effect, $\kappa = 10$

n_0	10	15	20	25	30	35
L_2 error (computed)	$4 \cdot 10^{-14}$	$2 \cdot 10^{-13}$	$1 \cdot 10^{-10}$	$2 \cdot 10^{-7}$	$8 \cdot 10^{-4}$	7.3
L_2 error (analytical)	$2 \cdot 10^{-14}$	$9 \cdot 10^{-13}$	$4 \cdot 10^{-10}$	$4 \cdot 10^{-7}$	$3 \cdot 10^{-3}$	40.5

Table 2

Degradation of the error due to the finite precision effect, $\kappa = 100$

n_0	100	110	120	130	140	150
L_2 error (computed)	$2 \cdot 10^{-13}$	$2 \cdot 10^{-12}$	$3 \cdot 10^{-10}$	$4 \cdot 10^{-7}$	10^{-3}	26.6
L_2 error (analytical)	10^{-13}	$3 \cdot 10^{-12}$	$9 \cdot 10^{-10}$	10^{-6}	$4 \cdot 10^{-3}$	43

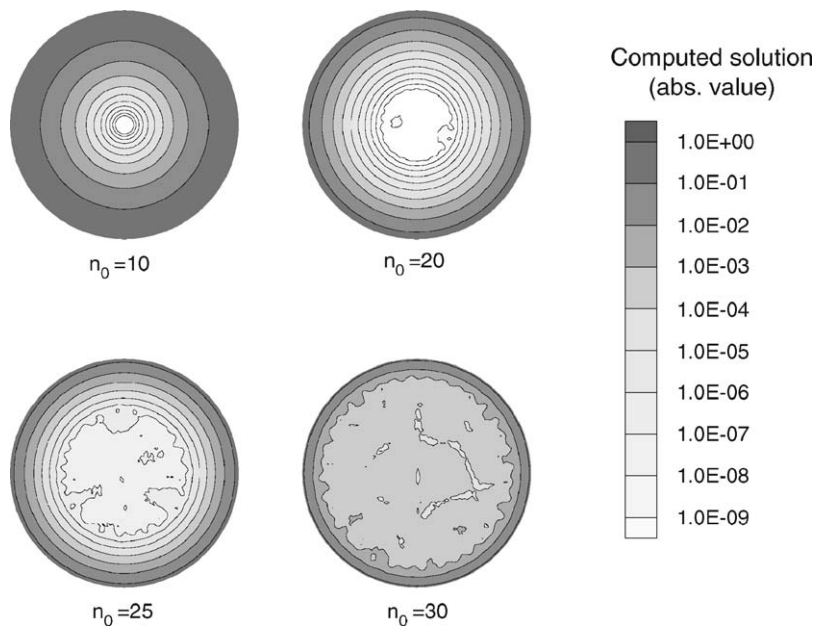


Fig. 3. Degradation of the plane wave approximation in the computational domain.

This simple numerical test reveals the difficulty for the plane wave basis to approximate the evanescent wave field located near the origin where it ‘superoscillates’ (this term is borrowed from Berry [4]). There is another practical situation where the plane wave basis is likely to break down: when attempting to recover a field emitted by a singular source very close to the computational domain. In this case, the boundary data g has high-order Fourier coefficients due to the $1/r^\alpha$ behavior in the vicinity of the source.

3. Elastic wave equations

We consider the propagation of waves in an elastic medium with Lamé constants μ, λ and density ρ . We restrict ourselves to the unit disc Ω by introducing the reduced wave numbers

$$\kappa_1 = \omega \frac{h}{2} \sqrt{\frac{\rho}{2\mu + \lambda}} \quad \text{and} \quad \kappa_2 = \omega \frac{h}{2} \sqrt{\frac{\rho}{\mu}},$$

where subscripts 1 and 2, respectively, refer to the pressure wave (P) and the shear wave (S). In the sequel, the P- and S-wave are referred to as the associated propagative plane wave. Let \mathbf{u} be the displacement field, the Dirichlet problem then writes

$$\mu \nabla^2 \mathbf{u} + (\lambda + \mu) \nabla \nabla \cdot \mathbf{u} + \rho \omega^2 \mathbf{u} = 0 \quad \text{on } \Omega, \quad (37)$$

$$\mathbf{u} = \mathbf{f} \quad \text{on } \gamma. \quad (38)$$

Here again, \mathbf{f} is assumed to be given by its Fourier series converging pointwise on $[0, 2\pi]$,

$$\mathbf{f}(\theta) = \sum_{n \in \mathbb{Z}} (\hat{f}_n^r \mathbf{e}_r + \hat{f}_n^\theta \mathbf{e}_\theta) e^{in\theta}, \quad (39)$$

where $(\mathbf{e}_r, \mathbf{e}_\theta)$ denotes the conventional polar basis.

3.1. Error analysis

Let us first introduce the Helmholtz decomposition for the field \mathbf{u} , namely

$$\mathbf{u} = \nabla \Phi_1 + \nabla^\perp \Phi_2, \quad (40)$$

where the Lamé potentials Φ_1 (resp. Φ_2) are solutions of the Helmholtz equation with wave number κ_1 (resp. κ_2) and thus admit the decomposition

$$\Phi_j(\mathbf{x}) = \sum_{n \in \mathbb{Z}} A_{j,n} J_n(\kappa_j r) e^{in\theta}, \quad j = 1, 2. \quad (41)$$

The decomposition is unique only if the 2×2 system

$$\begin{pmatrix} \kappa_1 J'_n(\kappa_1) & -in J_n(\kappa_2) \\ in J_n(\kappa_1) & \kappa_2 J'_n(\kappa_2) \end{pmatrix} \begin{pmatrix} A_{1,n} \\ A_{2,n} \end{pmatrix} = \begin{pmatrix} \hat{f}_n^r \\ \hat{f}_n^\theta \end{pmatrix} \quad (42)$$

is invertible for all $n \in \mathbb{Z}$ and this will always be assumed in the present discussion. Following the technique of the previous section, we split the Lamé potentials in three parts:

$$\begin{aligned} \nabla \Phi_1(\mathbf{x}) &= \sum_{q=1}^{Q_1} \nabla \Psi(\kappa_1; 2\pi q / Q_1, \mathbf{x}) \left(\frac{1}{Q_1} \sum_{|n| \leq N_1} A_{1,n} \frac{e^{in2\pi q / Q_1}}{i^n} \right) \\ &\quad + \sum_{|n| \leq N_1} \frac{A_{1,n}}{2\pi i^n} \nabla \varepsilon_{Q_1,n}(\kappa_1; \mathbf{x}) + \sum_{|n| > N_1} A_{1,n} \nabla (J_n(\kappa_1 r) e^{in\theta}). \end{aligned} \quad (43)$$

The first term clearly reveals a P-wave approximation of the solution using Q_1 directions evenly distributed over the unit circle. The remaining terms are the approximation error. Obviously, a similar decomposition holds for $\nabla^\perp \Phi_2$ with Q_2 S-waves.

Using recurrence relations (A.2), (A.3) and results established for the Helmholtz problem, it can be shown that the second term in (43) satisfies

$$\left\| \sum_{|n| \leq N_1} \frac{A_{1,n}}{2\pi i^n} \nabla \varepsilon_{Q_1,n}(\kappa_1; \mathbf{x}) \right\|_{L^\infty(\Omega)} < \frac{4\kappa_1 \beta_{N_1}^{Q_1}}{\beta_{N_1}^{Q_1} - 1} J_{Q_1-N_1-1}(\kappa_1) \sum_{|n| \leq N_1} |A_{1,n}|, \quad (44)$$

where the ratio $\beta_{N_1} = (Q_1 - N_1 - 1)/\kappa_1$ is strictly greater than one. Moreover, if $N_1 \geq \kappa_1 + 1$ then the following holds for the third term,

$$\left\| \sum_{|n| > N_1} A_{1,n} \nabla (J_n(\kappa_1 r) e^{in\theta}) \right\|_{L^\infty(\Omega)} \leq 2\kappa_1 \sum_{|n| > N_1} |A_{1,n}| J_{|n|-1}(\kappa_1). \quad (45)$$

By repeating the same operation for the potential Φ_2 , we can now state

Lemma 3. Define the system of P- and S-waves as

$$W_1(Q_1) = \text{span}\{\nabla \Psi(\kappa_1; \phi_q, \mathbf{x}), \phi_q = 2\pi q/Q_1, q = 1, \dots, Q_1\},$$

$$W_2(Q_2) = \text{span}\{\nabla^\perp \Psi(\kappa_2; \phi_q, \mathbf{x}), \phi_q = 2\pi q/Q_2, q = 1, \dots, Q_2\}.$$

Let the number of plane waves $Q_1 > 2(\kappa_1 + 1)$ and $Q_2 > 2(\kappa_2 + 1)$ be chosen such that the two sets $I_1 = [\kappa_1 + 1, Q_1 - \kappa_1 - 1] \cap \mathbb{N}$ and $I_2 = [\kappa_2 + 1, Q_2 - \kappa_2 - 1] \cap \mathbb{N}$ are not empty. Then, the best approximation in $W_1(Q_1) + W_2(Q_2)$ of the Dirichlet problem (37), (38) satisfies the inequality

$$\min_{\mathbf{w} \in W_1(Q_1) + W_2(Q_2)} \|\mathbf{u} - \mathbf{w}\|_{L^\infty(\Omega)} < \sum_{j=1,2} 2\kappa_j \min_{N_j \in I_j} \left\{ \sum_{|n| > N_j} |A_{j,n}| J_{|n|-1}(\kappa_j) + \frac{2\beta_{N_j}^{Q_j} J_{Q_j-N_j-1}(\kappa_j)}{\beta_{N_j}^{Q_j} - 1} \sum_{|n| \leq N_j} |A_{j,n}| \right\}, \quad (46)$$

where $\beta_{N_j} = (Q_j - N_j - 1)/\kappa_j$.

Obviously, in case the Fourier series f is finite then more precise estimates can be established along the line of Lemma 2.

In order to give a practical example of the previous lemma, let us assume the data f stems from the sum of a P- and a S-wave both travelling in arbitrary directions ϕ^1, ϕ^2 , so that the exact solution reads

$$\mathbf{u}_I(\mathbf{x}) = \frac{1}{\kappa_1} \nabla \Psi(\kappa_1; \phi^1, \mathbf{x}) + \frac{1}{\kappa_2} \nabla^\perp \Psi(\kappa_2; \phi^2, \mathbf{x}) \quad (47)$$

(amplitudes have been normalized to unity). In that case we have directly $|A_{j,n}| = 1/\kappa_j$. Moreover it is easy to see that, if $N_j > \kappa_j + 1$ then

$$\sum_{|n| > N_j} J_{|n|-1}(\kappa_j) < \frac{2J_{N_j-1}(\kappa_j)}{(N_j - 1)/\kappa_j - 1}.$$

As for the Helmholtz equation, a sharp estimate can be obtained by choosing $N_j = Q_j/2$ (the Q_j 's are assumed to be even for simplicity) giving

$$\min_{\mathbf{w} \in W_1(Q_1) + W_2(Q_2)} \|\mathbf{u}_I - \mathbf{w}\|_{L^\infty(\Omega)} < \sum_{j=1,2} 4J_{\beta_j \kappa_j}(\kappa_j) \left(\frac{1}{\beta_j - 1} + \frac{\beta_j^{Q_j}(Q_j + 1)}{\beta_j^{Q_j} - 1} \right), \quad (48)$$

where

$$\beta_j = (Q_j/2 - 1)/\kappa_j > 1.$$

This shows that any arbitrary superposition of a P- and a S-wave of unit amplitude can be approximated in $W_1(Q_1) + W_2(Q_2)$ with an error behaving like $Q_1 J_{Q_1/2-1}(\kappa_1) + Q_2 J_{Q_2/2-1}(\kappa_2)$. This simple observation can be used to find the optimal value for Q_1 and Q_2 as shown in the next section.

3.2. Numerical experiments

The system matrix arising either from least square or collocation formulation does not admit analytical diagonalization due to the coexistence of the separate scales κ_1 and κ_2 . Here, we shall simply check the error estimates (48) using a collocation formulation. As for the Helmholtz equation, we make sure that the determinant of (42) is not too small so that our analysis is not spoiled by the nonuniqueness problem. We seek an approximate solution of (37), (38) with the Dirichlet condition

$$f = \frac{1}{\kappa_1} \nabla \Psi(\kappa_1; \pi/Q_1, \mathbf{x}) + \frac{1}{\kappa_2} \nabla^\perp \Psi(\kappa_2; \pi/Q_2, \mathbf{x}), \quad \mathbf{x} \in \gamma$$

by considering the finite linear combination of P- and S-waves

$$\mathbf{w}_{Q_1, Q_2}(\mathbf{x}) = \frac{1}{\kappa_1} \sum_{q=1}^{Q_1} \nabla \Psi(\kappa_1; 2\pi q/Q_1, \mathbf{x}) + \frac{1}{\kappa_2} \sum_{q=1}^{Q_2} \nabla^\perp \Psi(\kappa_2; 2\pi q/Q_2, \mathbf{x}). \quad (49)$$

As for the choice of Q_1 and Q_2 , two strategies are tested: (i) we choose $Q_2 = \kappa_2/\kappa_1 Q_1$ (this strategy has been applied in [8]) (ii) given Q_1 , we choose the smallest Q_2 such that the minimum of the error bound (48) is almost reached (up to 1 significant digit in our application). Fig. 4 shows the evolution of the errors obtained as well as the theoretical estimates (48). When dealing with an elastic medium for which the ratio κ_2/κ_1 is moderate, the optimization algorithm is not needed. However, if the ratio is relatively high, the optimization clearly offers substantial savings (moreover it has positive effects on the conditioning). It is interesting to check if this applies in a more general case. For this purpose let the boundary conditions stemming from the wave field:

$$\mathbf{u}_I(\mathbf{x}) = \frac{1}{\kappa_1} \nabla Y_0(\kappa_1 |\mathbf{x} - \mathbf{x}_{(1)}|) + \frac{1}{\kappa_2} \nabla^\perp Y_0(\kappa_2 |\mathbf{x} - \mathbf{x}_{(2)}|),$$

where Y_0 is the Bessel function of the second kind of order 0. The source position of the pressure wave is chosen to be located at $\mathbf{x}_{(1)} = (2, 2)$ and the source position of the shear wave is chosen to be located at $\mathbf{x}_{(2)} = (2, -2)$. Plots of the approximated elastic wave field are shown in Fig. 5 for three increasing values of Q_1 whereas Q_2 is estimated from the optimization algorithm. As is clearly illustrated, the almost perfectly reconstructed field (c) ($E_Q = 5 \cdot 10^{-5}$) is reached at a very low cost and use of the strategy (i) would require more degrees of freedom to achieve the same accuracy.

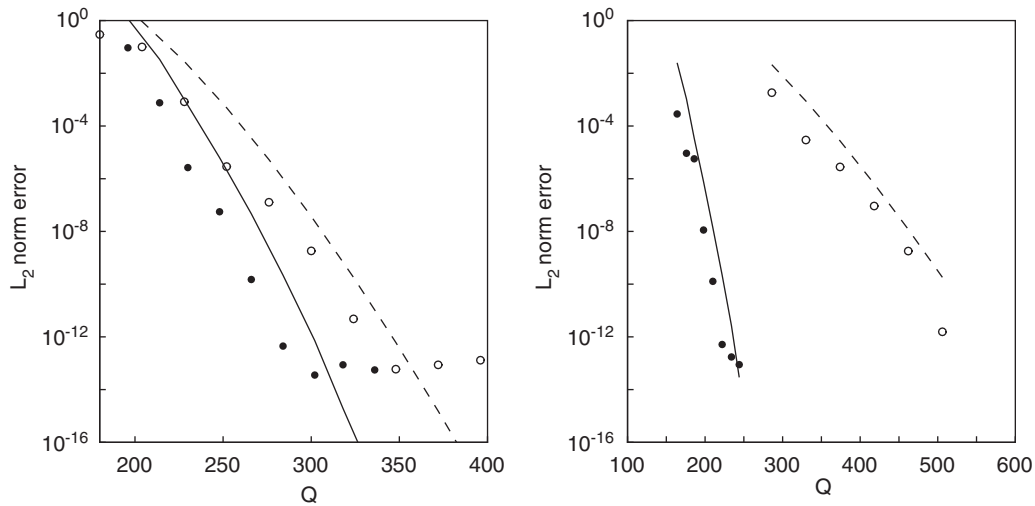


Fig. 4. Evolution of the error when approximating the sum of a P- and a S-wave ($Q = Q_1 + Q_2$). Left: $\kappa_1 = 25$ and $\kappa_2 = 50$. Right: $\kappa_1 = 5$ and $\kappa_2 = 50$. (i) $Q_2 = \kappa_2/\kappa_1 Q_1$: theoretical estimate (dashed line) and computed (hollow circle). (ii) Optimal: theoretical estimate (straight line) and computed (black circle).

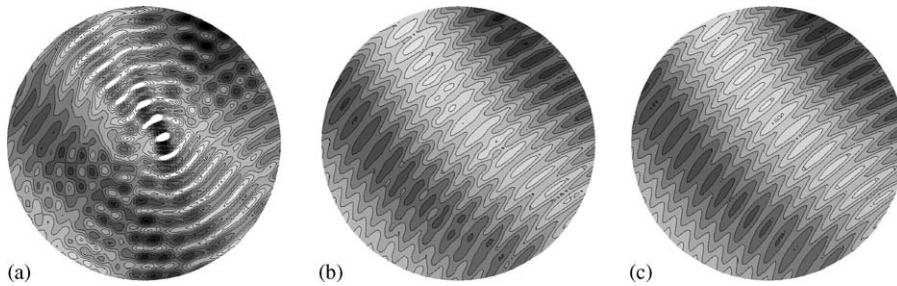


Fig. 5. P- and S-wave approximation of the elastic wave field in the computational domain (real part of the horizontal displacement). Isovalue interval: 0.05. $\kappa_1 = 5$ and $\kappa_2 = 50$ (a): $Q_1 = 12$, $Q_2 = 120$ ($E_Q = 0.16$). (b): $Q_1 = 20$, $Q_2 = 128$ ($E_Q = 6 \cdot 10^{-3}$). (c): $Q_1 = 24$, $Q_2 = 134$ ($E_Q = 5 \cdot 10^{-5}$).

4. Conclusion

Theoretical and numerical results presented in this paper highlight the numerical limitations of the plane wave basis due to the finite machine precision. This has some consequences for computational methods using plane wave basis functions. In practice, there should be a distinction between singular/evanescent regions (due to sources, boundary irregularities such as corners, abrupt changes of boundary conditions...) in which the wave field cannot properly be approximated with plane waves and regular/propagative regions for which the plane wave expansion is appropriate and appears to be the best basis from a computational point of view.

Appendix A. Properties of Bessel functions

We state here some useful properties of the Bessel function. These are given in [1, Chapter 9]. We call $J_\nu(x)$ the Bessel function of the first kind of order ν , then the following recurrence relations hold:

$$J'_\nu(x) = J_{\nu-1}(x) - \frac{\nu}{x} J_\nu(x), \quad (\text{A.1})$$

$$\frac{2\nu}{x} J_\nu(x) = J_{\nu-1}(x) + J_{\nu+1}(x), \quad (\text{A.2})$$

$$2J_\nu(x) = J'_\nu(x) - J_{\nu+1}(x). \quad (\text{A.3})$$

Now, if we restrict ourselves to real values x, y, ν such that

$$\nu \geq y \geq x > 0.$$

Then the following holds

$$J_\nu(x) > 0 \quad \text{and} \quad J'_\nu(x) > 0. \quad (\text{A.4})$$

A direct consequence of the second inequality is

$$J_\nu(y) \geq J_\nu(x) \quad (\text{A.5})$$

and the use of (A.1) gives (replace ν by $\nu + 1$)

$$\frac{J_{\nu+1}(x)}{J_\nu(x)} < \frac{x}{\nu+1} \quad \text{and consequently} \quad \frac{J_{\nu+p}(x)}{J_\nu(x)} < \left(\frac{x}{\nu}\right)^p \quad (\text{A.6})$$

for any strictly positive integer p . Negative orders can be handled when they are integers, and we have

$$|J_\nu(x)| = J_{|\nu|}(x) \quad \forall \nu \in \mathbb{Z} \quad |\nu| \geq x > 0. \quad (\text{A.7})$$

For large orders, we have the principal asymptotic form

$$J_\nu(x) \sim \sqrt{\frac{1}{2\pi\nu}} \left(\frac{ex}{2\nu}\right)^\nu \quad \text{as } \nu \rightarrow \infty \quad (\text{A.8})$$

we close this section with the Jacobi–Anger expansion for the plane wave

$$\Psi(\kappa; \phi, x) = e^{i\kappa r \cos(\theta-\phi)} = \sum_{m \in \mathbb{Z}} i^m J_m(\kappa r) e^{im(\theta-\phi)} \quad (\text{A.9})$$

which, by inversion, yields the Bessel's first integral identity,

$$J_n(\kappa r) e^{in\theta} = \frac{1}{2\pi i^n} \int_0^{2\pi} \Psi(\kappa; \phi, x) e^{in\phi} d\phi. \quad (\text{A.10})$$

Appendix B. Proofs of some inequalities

All the inequalities stated in this paper are direct consequences of properties given in Appendix A. We shall start with the main result of the paper.

B.1. Proof of (10)

The first step is to choose Q high enough so that $Q > |n| + \kappa$. This ensures that

$$kQ + n > \kappa \quad \forall k \geq 1 \quad \text{and} \quad -kQ - n > \kappa \quad \forall k \leq -1.$$

Thus for any point $x \in \Omega$

$$\begin{aligned} |\varepsilon_{Q,n}(\kappa; x)| &\leq 2\pi \sum_{k \in \mathbb{Z} \setminus \{0\}} |J_{n+kQ}(\kappa r)| \leq 2\pi \sum_{k \in \mathbb{Z} \setminus \{0\}} J_{|n+kQ|}(\kappa) \\ &\leq 2\pi \sum_{k \geq 1} (J_{kQ+|n|}(\kappa) + J_{kQ-|n|}(\kappa)) \\ &\leq 4\pi \sum_{k \geq 1} J_{kQ-|n|}(\kappa). \end{aligned}$$

Now, consider the range of indices $|n| \leq N$ and take $Q > N + \kappa$. Then by using (A.6), we have

$$\begin{aligned} |\varepsilon_{Q,n}(\kappa; x)| &\leq 4\pi \sum_{k \geq 1} J_{kQ-N}(\kappa) \\ &< 4\pi \sum_{k \geq 1} \left(\frac{\kappa}{Q-N} \right)^{(k-1)Q} J_{Q-N}(\kappa). \end{aligned}$$

We introduce the ratio $\beta_N = (Q - N)/\kappa$ and we finally obtain

$$|\varepsilon_{Q,n}(\kappa; x)| < \frac{4\pi\beta_N^Q}{\beta_N^Q - 1} J_{Q-N}(\kappa). \quad (\text{B.1})$$

B.2. Proof of (24)

In order to estimate a upper bound for the smallest singular value, we split the infinite sum as follows:

$$\sigma_k = J_k^2(\kappa) + J_{k-Q}^2(\kappa) + A_k + B_k \quad (1 \leq k \leq Q),$$

where

$$A_k = \sum_{l \geq 1} J_{k+lQ}^2(\kappa) \quad \text{and} \quad B_k = \sum_{l \leq -2} J_{k+lQ}^2(\kappa) = \sum_{l \geq 2} J_{lQ-k}^2(\kappa).$$

To make some progress, let $Q > 2\kappa$. In that case

$$\frac{J_{k+lQ}^2(\kappa)}{J_{k+Q}^2(\kappa)} < \left(\frac{\kappa}{k+Q} \right)^{2Q(l-1)} < \left(\frac{1}{2} \right)^{2Q(l-1)} \quad (l \geq 2).$$

Thus, (we consider Q even to ease the demonstration)

$$A_k < J_{k+Q}^2(\kappa) \sum_{l \geq 1} \left(\frac{1}{4Q} \right)^{l-1} < \frac{J_{k+Q}^2(\kappa)}{1 - \frac{1}{4Q}} < \frac{4}{3} J_{k+Q}^2(\kappa) < \frac{4}{3} J_{Q/2}^2(\kappa).$$

A Similar treatment applies for B_k and leads to $B_k < 4/3 J_{Q/2}^2(\kappa)$. So we conclude that

$$\min_{1 \leq k \leq Q} \sigma_k \leq \sigma_{Q/2} < 5 J_{Q/2}^2(\kappa). \quad (\text{B.2})$$

References

- [1] M. Abramovitz, I.A. Stegun, Handbook of Mathematical Functions, Applied Mathematical Series, 10th ed., National Bureau of Standards, US Government Printing Office, Washington, DC, 1972.
- [2] I. Babuška, J.M. Melenk, The partition of unity method, *Internat. J. Numer. Methods Eng.* 40 (1997) 727–758.
- [3] A.H. Barnett, Dissipation in Deforming Chaotic Billiards, Ph.D. Thesis, Department of Physics, Harvard University, 2000.
- [4] M.V. Berry, Evanescent and real waves in quantum billiards and Gaussian Beams, *J. Phys. A: Math. Gen.* 27 (1994) L391–L398.
- [5] Q. Carayol, F. Collino, Estimates in the fast multipole method for scattering problems Part 1: Truncation of the Jacobi-Anger series, *ESAIM: Math. Model. Numer. Anal.* 38 (2) (2004) 371–394.
- [6] O. Cessenat, B. Després, Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem, *SIAM J. Numer. Anal.* 35 (1) (1998) 255–299.
- [7] B. Després, Sur une formulation variationnelle de type ultra-faible, *C. R. Acad. Sci. Paris* 318 (1994) 939–944.
- [8] T. Huttunen, P. Monk, F. Collino, J.P. Kaipio, The ultra weak variational formulation for elastic wave problems, *SIAM J. Sci. Comp.* 25 (5) (2004) 1717–1742.
- [9] T. Huttunen, P. Monk, J.P. Kaipio, Computational aspects of the ultra weak variational formulation, *J. Comput. Phys.* 182 (2002) 27–46.
- [10] O. Laghrouche, P. Bettess, R.J. Astley, Modelling of short wave diffraction problems using approximating systems of plane waves, *Internat. J. Numer. Methods Eng.* 54 (2002) 1501–1533.
- [11] P. Monk, D.-Q. Wang, A least-squares method for the Helmholtz equation, *Comp. Methods Appl. Mech. Eng.* 175 (1999) 121–136.
- [12] P. Ortiz, E. Sanchez, An improved partition of unity finite element model for diffraction problems, *Internat. J. Numer. Methods Eng.* 50 (2001) 2727–2740.
- [13] E. Perrey-Debain, O. Laghrouche, P. Bettess, J. Trevelyan, Plane-wave basis finite elements and boundary elements for three dimensional wave scattering, *Philos. Trans. Roy. Soc. London A* 362 (2004) 629–645.